STATISTICS WORKSHEET-1

1. Bernoulli random variables take (only) the values 1 and 0.

Ans- True

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

Ans) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?

Ans - Modeling bounded count data

4. Point out the correct statement.

Ans- All of the mentioned

5. _____ random variables are used to model rates.

Ans- Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.

Ans- False

7. Which of the following testing is concerned with making decisions using data?

Ans- Hypothesis

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.

Ans - 10

9. Which of the following statement is incorrect with respect to outliers?

Ans - Outliers cannot conform to the regression relationship

10. What do you understand by the term Normal Distribution?

Ans - *A normal distribution is also called Gaussian distribution refers to a probability distribution where the values of a random variable are distributed symmetrically. These values are equally distributed on the left and the right side of the central tendency. Thus, a bell-shaped curve is formed.*

FORMULA: Y= 12π √e-(x-μ)22

Where -

Mμ=mean

Σ=standard deviation

Π=3.14159

E =2.71828

11. How do you handle missing data? What imputation techniques do you recommend?

Ans -Missing data appear when no value is available in one or more variables of an individual.

Deletions. Pairwise Deletion. Listwise Deletion/ Dropping rows. Dropping complete columns.

Basic Imputation Techniques. Imputation with a constant value. Imputation using the statistics (mean, median, mode)

K-Nearest Neighbor Imputation.

12. What is A/B testing?

Ans- An AB test is an example of statistical hypothesis testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

1**3. Is mean imputation of missing data acceptable practice?**

Ans-

14. What is linear regression in statistics?

Ans-In statistics, linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variable) denoted X. The case of one explanatory variable is called simple linear regression.

15. What are the various branches of statistics ?

Ans-The two branches of statistics are descriptive statistics and inferential statistics. All these branches of statistics follow a specific scientific approach which makes them equally essential to every statistic.